Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently amended) A method for generating a ~~representation~~ fingerprint of a document comprising:

obtaining a plurality of overlapping blocks by sampling the document;

generating a set of checksum values for the plurality of overlapping blocks;

choosing a subset of the ~~plurality of overlapping blocks~~ set of checksum values, where the subset is less than an entirety of the ~~plurality of overlapping blocks~~ set of checksum values; and

compacting the subset of the ~~plurality of overlapping blocks~~ set of checksum values to obtain ~~the representation~~ the fingerprint of the document by addressing bits in the fingerprint with particular values and flipping bits in the fingerprint, where a number of times a particular bit is flipped is based on a number of checksum values in the subset that correspond to the particular value addressed to the particular bit.

2. (Canceled)

3. (Currently amended) The method of claim 1, ~~wherein~~ where the ~~representation of the document includes a~~ fingerprint [[of]] comprises a predetermined length.

4. (Currently amended) The method of claim 3, ~~wherein~~ where the predetermined length is eight or sixteen bytes.

5. (Currently amended) The method of claim 1, further comprising:

~~generating checksum values for the plurality of overlapping blocks~~

initializing the fingerprint before flipping bits in the fingerprint.

6. (Currently amended) The method of claim 5, ~~wherein~~ where choosing a subset of the ~~plurality of overlapping blocks~~ set of checksum values includes selecting a predetermined number of the smallest checksum values.

7. (Currently amended) The method of claim 5, ~~wherein~~ where choosing a subset of the ~~plurality of overlapping blocks~~ set of checksum values includes selecting a predetermined number of the largest checksum values.

8. (Currently amended) The method of claim 1, further comprising:

hashing the subset of the ~~plurality of overlapping blocks~~ set of checksum values to a length for ~~indexing the representation of the document~~ addressing the bits of the fingerprint.

9. (Currently amended) The method of claim 8, ~~wherein~~ where hashing the subset of the ~~plurality of overlapping blocks~~ set of checksum values includes taking a

number of least significant bits of the subset of the ~~plurality of overlapping blocks~~ set of checksum values.

10. (Canceled)

11. (Currently amended) The method of claim 1, ~~wherein~~ where each of the plurality of overlapping blocks is of a predetermined length.

12. (Currently amended) The method of claim 11, ~~wherein~~ where obtaining a plurality of overlapping blocks further includes:

padding null characters to the document when a length of the document is below the predetermined length.

13. (Currently amended) A method for generating a ~~representation~~ fingerprint of a document comprising:

sampling the document to obtain a plurality of overlapping samples;

generating a set of checksum values for the plurality of overlapping samples;

selecting a predetermined number of the ~~plurality of overlapping samples~~ set of checksum values as those of the ~~samples~~ checksum values corresponding to a predetermined number of smallest ~~samples~~ checksum values or a predetermined number of largest ~~samples~~ checksum values; and

setting bits in the ~~representation~~ fingerprint of the document ~~based on the selected predetermined number of the samples~~ by addressing bits in the fingerprint with particular

values and flipping bits in the fingerprint, where a number of times a particular bit is flipped is based on a number of checksum values in the subset that correspond to the particular value addressed to the particular bit.

14. (Currently amended) The method of claim 13, ~~wherein~~ where the ~~representation of the document includes a~~ fingerprint [[of]] comprises a predetermined length.

15. (Currently amended) The method of claim 14, ~~wherein~~ where the predetermined length is eight or sixteen bytes.

16. (Canceled)

17. (Currently amended) The method of claim 13, further comprising:

hashing the predetermined number of the ~~samples~~ checksum values to a length for indexing the ~~representation~~ fingerprint of the document.

18. (Currently amended) The method of claim 17, ~~wherein~~ where hashing the predetermined number of the ~~samples~~ checksum values includes taking a number of least significant bits of the predetermined number of ~~samples~~ checksum values.

19. (Currently amended) The method of claim 17, ~~wherein~~ where setting bits in the ~~representation~~ fingerprint of the document includes flipping a bit in the

~~representation~~ fingerprint of the document when the particular value addressed to the bit ~~is addressed by the~~ corresponds to a particular one of the hashed ~~samples~~ checksum values.

20. (Currently amended) A computer-implemented device comprising:

a memory to store instructions for implementing:

a fingerprint creation component to generate a fingerprint of a predetermined length for an input document, the fingerprint generated by

sampling the input document to obtain samples,

generating a set of checksum values for the samples;

choosing a subset of the ~~samples~~ set of checksum values, where the subset is less than an entirety of the ~~samples~~ set of checksum values, and

generating the fingerprint from the subset of the ~~samples~~ set of checksum values by ~~compacting the subset of the samples~~ addressing bits in the fingerprint with particular values and flipping bits in the fingerprint, where a number of times a particular bit is flipped is based on a number of checksum values in the subset that correspond to the particular value addressed to the particular bit, and

a similarity detection component to compare pairs of fingerprints to determine whether the pairs of fingerprints correspond to near-duplicate documents; and

a processor to execute the instructions in the memory.

21. (Previously presented) The computer-implemented device of claim 20, further including:

a search engine to return documents to a user as a single link when the documents are determined to correspond to near-duplicate documents.

22. (Currently amended) The computer-implemented device of claim 20, ~~wherein~~ where the similarity detection component compares the pairs of fingerprints by calculating a hamming distance.

23. (Currently amended) The computer-implemented device of claim 22, ~~wherein~~ where the similarity detection component determines that the pairs of documents correspond to near-duplicate documents when the hamming distance is below a threshold.

24. (Currently amended) The computer-implemented device of claim 20, ~~wherein~~ where the fingerprint creation component additionally:

chooses the subset as a predetermined number of smallest checksums calculated from the subset of the samples.

25. (Currently amended) The computer-implemented device of claim 20, ~~wherein~~ where the fingerprint creation component additionally:

chooses the subset as a predetermined number of largest checksums calculated from the subset of the samples.

26. (Currently amended) A computer-implemented device comprising:

memory to store instructions for implementing:

means for sampling a document to obtain a plurality of overlapping blocks,

means for generating a set of checksum values for the plurality of overlapping blocks;

means for choosing a subset of the ~~plurality of overlapping blocks~~ set of checksum values, where the subset is less than an entirety of the ~~plurality of overlapping blocks~~ set of checksum values, and

means for compacting the subset of the ~~plurality of overlapping blocks~~ set of checksum values to obtain a ~~compact representation of the document~~ a fingerprint of the document by addressing bits in the fingerprint with particular values and flipping bits in the fingerprint, where a number of times a particular bit is flipped is based on a number of checksum values in the subset that correspond to the particular value addressed to the particular bit; and

a processor to execute the instructions in memory.

27. (Canceled)

28. (Currently amended) The computer-implemented device of claim [[27]] 26, ~~wherein~~ where the means for choosing a subset of the ~~plurality of overlapping blocks~~ set of checksum values chooses the subset as a predetermined number of the smallest checksum values.

29. (Currently amended) The computer-implemented device of claim [[27]] 26, ~~wherein~~ where the means for choosing a subset of the ~~plurality of overlapping blocks~~ set of checksum values chooses the subset as a predetermined number of the largest checksum values.

30. (Currently amended) The computer-implemented device of claim [[27]] 26, ~~wherein~~ where the means for compacting the subset of ~~plurality of overlapping blocks~~ the set of checksum values includes means for ~~flipping bits in the compact representation that are addressed by a hashed version of~~ hashing the checksum values.

31. (Currently amended) A computer-readable memory device containing program instructions that, when executed by a processor, cause the processor to:

sample a document to obtain a plurality of overlapping samples;

generate a set of checksum values for the plurality of overlapping samples;

select a predetermined number of the ~~plurality of overlapping samples~~ set of checksum values as those of the ~~samples~~ checksum values corresponding to a predetermined number of smallest ~~samples~~ checksum values or a predetermined number of largest ~~samples~~ checksum values; and

set bits in a ~~representation~~ fingerprint of the document ~~based on the selected predetermined number of the samples~~ by addressing bits in the fingerprint with particular values and flipping bits in the fingerprint, where a number of times a particular bit is flipped is based on a number of checksum values in the subset that correspond to the particular value addressed to the particular bit.

32. (Currently amended) The computer-readable memory device of claim 31, further including program instructions that, when executed by the processor, cause the processor to:

hash the predetermined number of the ~~samples~~ checksum values to a length for ~~indexing the representation~~ setting the bits in the fingerprint of the document.

33. (Currently amended) The computer-readable memory device of claim 32, ~~wherein~~ where hashing the predetermined number of the ~~samples~~ checksum values includes taking a number of least significant bits of the predetermined number of ~~samples~~ checksum values.

34. (Currently amended) A computer-implemented device comprising:

memory to store instructions for implementing:

means for sampling a document to obtain a plurality of overlapping blocks,

means for calculating checksum values for the plurality of overlapping blocks,

means for choosing a subset of the ~~plurality of overlapping blocks based on the~~ calculated checksum values, where the subset is less than an entirety of the ~~overlapping blocks~~ calculated checksum values,

means for hashing the checksum values in the subset to a predetermined size; and

means for setting bits in a ~~compact representation~~ fingerprint of the document ~~based on the subset of the plurality of overlapping blocks for~~ by flipping bits in the ~~compact representation~~ fingerprint, where a particular bit in the fingerprint is addressed with a particular hashed checksum value, and where a number of times the particular bit in the fingerprint is flipped corresponds to a number of times the particular hashed checksum value occurs in the subset ~~that are addressed by a hashed version of the checksum values~~; and

a processor to execute the instructions in memory.

35. (New) The computer-implemented device of claim 34, further comprising:

means for initializing the fingerprint before setting bits in the fingerprint.

36. (New) The computer-implemented device of claim 34, further comprising:

means for comparing pairs of fingerprints to determine whether the pairs of fingerprints correspond to near-duplicate documents.